

# UCSF

## UC San Francisco Previously Published Works

### Title

Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways.

### Permalink

<https://escholarship.org/uc/item/0gq9h2m6>

### Journal

Nature immunology, 20(7)

### ISSN

1529-2908

### Authors

Der, Evan  
Suryawanshi, Hemant  
Morozov, Pavel  
et al.

### Publication Date

2019-07-01

### DOI

10.1038/s41590-019-0386-1

Peer reviewed



Published in final edited form as:

*Nat Immunol.* 2019 July ; 20(7): 915–927. doi:10.1038/s41590-019-0386-1.

## Tubular Cell and Keratinocyte Single-cell Transcriptomics Applied to Lupus Nephritis Reveal Type I IFN and Fibrosis Relevant Pathways

Evan Der<sup>#,1</sup>, Hemant Suryawanshi<sup>#,2</sup>, Pavel Morozov<sup>2</sup>, Manjunath Kustagi<sup>2</sup>, Beatrice Goilav<sup>3</sup>, Saritha Ranabathou<sup>4</sup>, Peter Izmirly<sup>5</sup>, Robert Clancy<sup>5</sup>, H. Michael Belmont<sup>5</sup>, Mordecai Koenigsberg<sup>6</sup>, Michele Mokrzycki<sup>7</sup>, Helen Rominieki<sup>8</sup>, Jay A. Graham<sup>8</sup>, Juan P. Rocca<sup>8</sup>, Nicole Bornkamp<sup>5</sup>, Nicole Jordan<sup>1</sup>, Emma Schulte<sup>1</sup>, Ming Wu<sup>5</sup>, James Pullman<sup>9</sup>, Kamil Slowikowski<sup>10</sup>, Soumya Raychaudhuri<sup>10</sup>, Joel Guthridge<sup>11</sup>, Judith James<sup>11</sup>, Jill Buyon<sup>\*,5</sup>, Thomas Tuschl<sup>\*,2</sup>, Chaim Putterman<sup>\*,1</sup>, and Accelerating Medicines Partnership in RA/SLE network<sup>12</sup>

<sup>1</sup>Albert Einstein College of Medicine, Bronx, New York, USA.

<sup>2</sup>Laboratory for RNA Molecular Biology, The Rockefeller University, New York, New York, USA.

<sup>3</sup>Pediatric Nephrology, The Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, New York, USA.

<sup>4</sup>Pediatric Nephrology, Arkansas Children's Hospital, University of Arkansas Medical Sciences, Little Rock, Arkansas, USA

<sup>5</sup>New York University School of Medicine, New York, New York, USA.

<sup>6</sup>Department of Radiology, Montefiore Medical Center, Bronx, New York, USA.

<sup>7</sup>Division of Nephrology, Department of Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, New York, USA.

<sup>8</sup>Montefiore Einstein Center for Transplantation, Montefiore Medical Center, Bronx, New York, USA.

<sup>9</sup>Clinical Pathology, Montefiore Medical Center, Bronx, New York, USA.

<sup>10</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

<sup>11</sup>Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Co-corresponding authors: [chaim.putterman@einstein.yu.edu](mailto:chaim.putterman@einstein.yu.edu), [ttuschl@mail.rockefeller.edu](mailto:ttuschl@mail.rockefeller.edu), [jill.buyon@nyumc.org](mailto:jill.buyon@nyumc.org).

Author contributions

JB, TT, and CP conceived the study with help from S. Ranabathu, JJ, J. Guthridge and S. Raychaudhuri. Input regarding the skin came from RC and HMB. ED, HS, and S. Ranabathu performed all biopsy dissociations and single-cell experiments. BG, PI, HMB, M. Koenigsberg, M. Mokrzycki, NJ, NB, and ES assisted with patient consent and sample acquisition of LN biopsies. HR, JR, J. Graham assisted with patient consent and sample acquisition of live kidney donor tissue. Renal biopsy histology was evaluated by MW and JP. HMB and PI performed all skin biopsies. Analysis was performed by ED, HS, PM, KS, and M. Kustagi. ED, JB, TT and CP prepared and wrote the manuscript.

<sup>#</sup>Authors contributed equally

Competing interests

The authors declare no competing interests.

<sup>12</sup>Lupus investigators in the Accelerating Medicines Partnership in RA/SLE network include the following individuals: Jennifer Anolik, William Apruzzese, Arnon Arazi, Celine Berthier, Michael Brenner, Jill Buyon, Robert Clancy, Sean Connery, Melissa Cunningham, Maria Dall'Era, Anne Davidson, Evan Der, Andrea Fava, Chamith Fonseka, Richard Furie, Dan Goldman, Rohit Gupta, Joel Guthridge, Nir Hacohen, David Hildeman, Paul Hoover, Raymond Hsu, Judith James, Ruba Kado, Ken Kalunian, Diane Kamen, Mattias Kretzler, Holden Maecker, Elena Massarotti, William McCune, Maureen McMahon, Meyeon Park, Fernanda Payan-Schober, William Pendergraft, Michelle Petri, Mina Pichavant, Chaim Putterman, Deepak Rao, Soumya Raychaudhuri, Kamil Slowikowski, Hemant Suryawanshi, Thomas Tuschl, PJ Utz, Dia Waguespack, David Wofsy, Fan Zhang

## Abstract

The molecular and cellular processes that lead to renal damage and to the heterogeneity of lupus nephritis (LN) are not well understood. We applied single-cell RNA sequencing (scRNA-seq) to renal biopsies from patients with LN and evaluated skin biopsies as a potential source of diagnostic and prognostic markers of renal disease. Type I interferon (IFN) response signatures in tubular cells and in keratinocytes distinguished patients with LN from healthy control subjects. Moreover, a high IFN response signature and fibrotic signature in tubular cells were each associated with failure to respond to treatment. Analysis of tubular cells from patients with proliferative, membranous, and mixed LN indicated pathways relevant to inflammation and fibrosis, which offer insight into their histological differences. In summary, we applied scRNA-seq to LN to deconstruct its heterogeneity and identify novel targets for personalized approaches to therapy.

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease that can affect multiple organs including the heart, brain, skin, lungs, and kidneys. SLE is characterized by the production of autoreactive antibodies against nuclear antigens such as ribonucleoproteins, dsDNA, and histones<sup>1</sup>. Lupus nephritis (LN) affects ~50% of patients with SLE and is a major contributor to mortality and morbidity<sup>2</sup>. Although the exact pathogenesis has yet to be fully characterized, immune complex deposition in and along the glomerular basement membrane and in the mesangial matrix, with secondary inflammation and proliferation of mesangial and endothelial cells, are hallmarks of the disease. Additionally, hypercellularity of mesangial and endothelial cells, as well as interstitial and glomerular fibrosis, are common features of chronicity and disease progression.

These immune, inflammatory, and parenchymal cell proliferative responses of LN have visible and heterogeneous histopathologic manifestations, which can be monitored by renal biopsy and evaluated according to the International Society of Nephrology/Renal Pathology Society (ISN/RPS) 2003 Lupus Nephritis Classification System<sup>3</sup>. The spectrum of glomerular pathology is variable not only between patients, but frequently within the same patient. Moreover, neither initial clinical manifestations nor treatment responses uniformly correlate with the histologic class of glomerular injury. Thus, clinical findings and biopsy alone are insufficient for accurate prognosis and further measures need to be developed to improve treatment and prognostic decisions. Additionally, the molecular basis for the observed histopathology is not yet fully characterized and further heterogeneity may exist,

which could explain the difficulty in accurately predicting response to treatment. For instance, fibrosis has been associated with poor response to treatment, but the underlying mechanisms initiating and promoting fibrosis are not fully understood. A further limitation within the ISN/RPS classification system is that histologic analysis is completely based on glomerular changes, despite a growing body of literature suggesting that the tubulointerstitial space is more predictive of response to therapy and prognosis, with infiltrates and fibrosis associated with poor renal outcome<sup>4-6</sup>.

Other potential and more accessible tissue sites than the kidney could also be exploited to obtain tissue for biomarkers of SLE progression<sup>7</sup>. Discovery of signatures in readily accessible tissue such as the skin, which even in non-lesional areas can have immunoglobulin deposition at the dermoepidermal junction (referred to as the lupus band test) analogous to that seen in the kidney<sup>8</sup>, would greatly facilitate early diagnosis and treatment decisions in a much less invasive manner. A previous study demonstrated an interferon signature in the keratinocytes from biopsies of non-lesional non-sun exposed skin of patients with LN compared to healthy control subjects<sup>9</sup>. This provides a rationale for using skin as a potential surrogate of renal disease, which could be sampled serially to follow response.

Single-cell RNA-sequencing (scRNA-seq) is a transcriptomic technology resolving cell type contributions in tissues<sup>10,11</sup>. This technique has been applied to a number of complex renal diseases including renal cell carcinoma<sup>12,13</sup> as well as to LN<sup>9</sup>. When resolved at a cell type level, transcriptome analysis yields valuable information regarding intercellular signaling responses and cell-type-specific pathways involved in promoting and maintaining LN. Here, we applied scRNA-seq to renal biopsies of patients with LN to identify novel clinically relevant prognostic markers, uncover intercellular interactions, and elucidate key pathways underlying the histological classes of LN.

## Results

### Samples and data acquisition

A total of 21 renal tissue samples were collected from patients with LN undergoing a clinically indicated renal biopsy (Supplementary Table 1). Of these patients, 17 also had a skin punch biopsy performed at the time of the renal biopsy. In addition to patients with LN, 3 biopsy pairs of control skin and renal tissue were obtained from healthy control subjects undergoing a nephrectomy for kidney transplant donation. Cell suspensions from skin and kidney biopsies of the same patient were loaded into separate compartments present on a single chip capturing about 250 cells per tissue type (Fig. 1a). The cells captured per chip were sequenced at an approximate depth of 200,000 reads/cell disregarding calibrator spike reads. A total of 19,200 wells were sequenced; however, only data originating from 6,041 wells confirmed by microscopy to contain single cells and resulting in a minimum read count of 10,000 were retained for downstream bioinformatics analysis.

## Cell lineage determination

Cell lineage determination was done using a principal component analysis (PCA) and a graph-based clustering approach. In an iterative process, we filtered out cells of abnormally high or low gene counts indicating doublet cell captures or poor-quality cells, respectively, resulting in 4,019 cells entering the final analysis. While mitochondrial cutoffs have been employed to set aside dead or dying cells, mitochondrial percentage cutoffs were initially developed in blood cells where there is little variation in mitochondrial content across cell types. In epithelial cells, we and others have observed mitochondrial content to vary widely across cell types, and mitochondrial percent cut-offs suitable for hematopoietic cells were deemed inaccurate to determine dead or dying non-immune cells<sup>14</sup>. Dispersion and mean expression values were calculated for each gene to identify highly variable genes, which were subjected to PCA, and resulted in 11 significant principal components. Graph-based clustering followed by t-Distributed Stochastic Neighbor Embedding (tSNE) was used to collapse the principal components into two dimensions and resulted in 6 distinct clusters of cell types (Fig. 1b). Differential expression analysis identified mutually exclusive sets of genes, which were characteristic of the cell lineage and frequently included established markers of particular cell types. The top 30 most differentially expressed markers in each cluster are provided in Supplementary Table 2. For example, tubular cells uniquely expressed *UMOD* and *SLC12A1*, whereas keratinocytes uniquely expressed *KRT1* and *KRT10*. Fibroblasts expressed many genes encoding extracellular matrix proteins including *DCN*, whereas endothelial cells distinctly expressed *FLT1* and *PECAMI*. Leukocytes expressed distinct myeloid, T cell, and B cell genes (*CD14*, *CD3G*, and *MZB1*, respectively) yet appeared as one cell type by tSNE analysis (Fig. 1c–e). Although we did not capture all known types of glomerular cell types, mesangial cells were recovered as indicated by high expression of their unique marker *TAGLN*. As anticipated, skin and renal biopsies were predominantly keratinocytes and tubular cells, respectively. The residual cell types represented smaller percentages and their relative abundance varied widely across samples, which is also a feature of clinical biopsies examined by standard diagnostic techniques (Fig. 1c). The relative contribution of each patient to each cluster is reported in Supplementary Table 3, and the patient identifier for each cell in the tSNE plot is depicted in Supplementary Fig. 1. The relative numerical representation of cell types within the tissue was also estimated by counting cells from the hematoxylin and eosin stained slides used for clinical diagnosis, and yielded similar percentages to those identified by PCA and tSNE analysis (Supplementary Fig. 2). None of these cell type percentages from either counting method correlated with clinical data obtained including activity, chronicity, or biopsy class (data not shown).

When averaged together across all renal cells, scRNA-seq expression resembled a bulk polyA-mRNA sequenced renal biopsy. Similarly, averaged skin single cells correlated with a bulk polyA-mRNA sequenced dissociated skin sample. Although averaged renal single cells also correlated with bulk sequenced skin and vice versa, they did so to a lesser extent than their originating tissue type (Supplementary Fig. 3).

When the keratinocyte subset identified by the first level of clustering analysis was once more subjected to clustering analysis, the presence of a small number of sweat gland cells

and melanocytes defined by expression of *DCD* and *MLANA*, respectively, became apparent (Fig. 2a–d)<sup>15,16</sup>. These cell types were excluded from participation in downstream comparative keratinocyte analysis. Similarly, the group of tubular cells identified by first level analysis was composed of various subtypes representative of the distinct nephron segments as previously reported (Fig. 3a–d)<sup>9</sup>. In summary, using graph-based clustering of the single-cell transcriptomes we were able to assign lineages to each cell and identify the cell types present in the biopsy samples, including both the major epithelial cell populations as well as rarer cell populations and cell subtypes.

### **LN skin and kidney epithelium indicate upregulation of type-I IFN response pathway genes**

It has been shown that type-I IFNs are elevated in both human and murine SLE and may drive pathogenesis, as IFN levels have been associated with disease flares in LN<sup>17</sup>. It was previously demonstrated in a small cohort of patients that keratinocytes from patients with LN show upregulation of IFN responsive genes compared to healthy control subjects<sup>9</sup>. Here, through cumulative distribution function analysis we confirmed this observation in a separate and larger cohort of patients and further expanded this finding to tubular cells (Fig. 4a). Type-I IFN response genes in tubular cells ( $p = 1.4e-11$ ) and keratinocytes ( $p = 3.3e-10$ ) from patients with LN were significantly higher expressed than those of healthy control subjects as indicated by the right-shifted curve of established IFN response genes compared to ubiquitously expressed genes (Fig. 4a–b). Using the tubular expression of IFN response genes we created an IFN response score for each patient, and found that patients who did not respond to treatment had significantly higher ( $p = 0.04$ ) IFN response scores compared to those who were either partial (50% reduction in proteinuria at 6 months post biopsy) or complete responders (urine protein-to-creatinine ratio (UPCR) mg/mg < 0.5) (Fig. 4c). The IFN response score in keratinocytes significantly correlated with the IFN response score in tubular cells ( $r = 0.61$ ,  $p = 0.004$ ; Fig. 4d). In contrast, correlations of clinical parameters including UPCR at the time of biopsy ( $p = 0.3$ ), Chronicity Index scores ( $p = 0.83$ ), and Activity Index scores ( $p = 0.14$ ) with both IFN response scores (Fig. 4e) and clinical response status (data not shown) were not significant.

To determine whether scRNA-seq data from kidney biopsy tissue provides more valuable information than that obtained by non-invasive means, scRNA-seq analysis was performed on peripheral blood mononuclear cells (PBMCs) collected at the time of biopsy, specifically 4 responders and 4 non-responders. The IFN response scores did not vary as a function of the clinical response in any of the identified cell populations including T cells, B cells, monocytes, and natural killer (NK) cells (data not shown). In conclusion, IFN response scores were elevated in tubular cells and keratinocytes from patients with LN patients compared to healthy control subjects, and elevated in tubular cells from patients with LN who did not respond to treatment compared to those with favorable therapeutic responses.

### **Patients non-responsive to treatment demonstrate higher expression of fibrotic extracellular matrix proteins as compared to responders**

To explore pathways other than those reflective of IFN signaling in patients who did not respond to therapy, differential expression analysis was performed on the average tubular cell profiles created for each patient. This analysis identified 301 significantly ( $p < 0.05$ )

differentially regulated genes (Fig. 5a). Enrichment analysis revealed significant ( $p < 0.001$ ) upregulation of genes encoding extracellular matrix (ECM) proteins and ECM-interaction proteins, reflective of an active fibrotic pathway in patients that were unresponsive to therapy compared to those who responded. A similar expression pattern has been previously reported and associated with tubular epithelial-mesenchymal transition (EMT) or tubular epithelial hypertrophy, both of which may have implications for progressive renal tubulointerstitial fibrosis<sup>18–21</sup>. Relevant to LN, tubulointerstitial fibrosis is a marker of poor prognosis<sup>5,6</sup> further supporting the finding of this expression in non-responders. Of clinical relevance, this gene signature may be predictive of a fibrotic response before it is measurable by standard histopathological assessment since the biopsies of some of these patients did not demonstrate fibrosis by typical scoring of tubulointerstitial damage. While it is acknowledged that ECM proteins are typically expressed by canonical fibroblasts, the cellular subset in this analysis expressed tubular cell markers, but not fibroblast markers such as *VIM* and *FSP1* (Fig. 5a), supporting that this observation was not simply due to fibroblast contamination. Finally, although it is possible that fibroblasts may also play an important role in the fibrotic pathways leading to tubulointerstitial fibrosis and progressive renal insufficiency in LN, too few fibroblasts were captured to assess any potential differences in the contribution of fibroblasts between groups (data not shown).

Two of the differentially expressed genes identified by pathway analysis as ECM interacting proteins, *TIMP1* and *SERPING*, that were upregulated in tubular cells of patients who did not respond to treatment, have previously been shown to be pro-fibrotic and associated with renal fibrosis<sup>22,23</sup>. Similarly, upregulation of the complement and coagulation cascades including *C1S* and *C1R* were also noted in non-responders (Fig. 5b)<sup>24</sup>.

A similar analysis was applied to the keratinocytes of non-responders and responders to assess the possibility of monitoring pathways activated in the epithelium of a tissue distant from the site of inflammation. Pathway enrichment analysis on the differentially expressed genes from keratinocytes of patients who did not respond to treatment also demonstrated upregulation of extracellular matrix (Fig. 5b). These included some overlapping and non-overlapping collagen genes expressed in the tubular cells; for instance, *COL1A1* was upregulated in both keratinocytes and tubular cells, but *COL17A1* was only upregulated in the keratinocytes of non-responders and not the tubular cells. The full list of differentially expressed genes from each comparison can be found in Supplementary Table 4.

Using logistic regression analysis on fibrotic genes in the tubular cells, an equation predicting response to treatment at 6 months post biopsy was created using genes identified as fibrotic markers among the differentially expressed genes between responders and non-responders. Four genes, *COL1A1*, *COL14A1*, *COL1A2*, and *COL5A2* were found to significantly explain variance and predict response to treatment with a 92% accuracy and an area under the curve (AUC) of 0.96 (Fig. 5c). We compared this signature to previously published 3-gene (*IL1RAP*, *NCAM1*, *FCAR*) and 5-gene (*IL1RAP*, *NCAM1*, *FCAR*, *IL28B*, and *C7*) signatures predicting response to treatment. We found that although these signatures were associated with response to treatment in our cohort, they were much less robust than the collagen-based signature we identified here with AUCs of 0.8 and 0.81<sup>25</sup> (Supplementary Fig. 4). Similar to the IFN response score, we found that the fibrosis score



in keratinocytes correlated with the same score in tubular cells ( $r = 0.45$ ,  $p = 0.04$ ; Fig. 5d left panel). Correlations between response to treatment and patient demographics (race, ethnicity, etc.) were explored, but none were found (data not shown). Similarly, correlations between tubular fibrosis scores and UPCR at the time of biopsy ( $p = 0.94$ ), Chronicity Index scores ( $p = 0.22$ ), and Activity Index scores ( $p = 0.07$ ) were not significant (Fig. 5d). Furthermore, there was no correlation between fibrosis seen by light microscopy and fibrosis scores (data not shown). PBMC populations including T cells, B cells, NK cells, and monocytes from the same patients showed no expression of extracellular matrix proteins or fibrotic markers by scRNA-seq (data not shown). This analysis demonstrated that patients who did not respond to treatment had significantly elevated levels of ECM proteins which through logistic regression were distilled to a 4-gene signature which could predict response to treatment at six months after biopsy.

### **Fibrotic pathways in kidney may be initiated by infiltrating cell receptor-ligand interactions**

Understanding the intercellular networks of communication can help elucidate potential targets for therapy in a cell-type-specific manner. scRNA-seq provides a unique starting point for deciphering ligand-receptor interactions by resolving gene expression according to cell type. Potential engagement of the highest expressed cognate receptors and ligands of cell types present in LN skin and kidney are indicated (Fig. 6). Many cells in the kidney including tubular cells expressed genes encoding various FGF receptors (FGFRs), such as *FGFR3*, at high levels. FGFs and FGFRs have been implicated in fibrosis in many organs including the kidney<sup>26</sup>. While it has been reported that FGF can be produced by epithelium<sup>27</sup>, in this study *FGF13* was expressed at high levels by infiltrating leukocytes, but not other renal cell types. Additionally, tubular cells expressed high levels of the gene encoding chemokine *CCL17* whose receptor *CCR1* was expressed within the leukocyte population, indicating a potential mechanism for attracting circulating leukocytes into the interstitium. Tubular cells also expressed high levels of *TNFSF10*, potentially signaling to leukocytes which expressed the gene encoding its receptor, *TNFRSF10A*. Similarly, keratinocytes expressed high levels of *FGFR3*, although in the skin the ligand was expressed in fibroblasts rather than leukocytes as in the kidney. By examining the cells within the leukocyte populations expressing these receptors and ligands and comparing their transcriptomic profiles to canonical markers we could further resolve the likely cell types of origin, and conclude that *FGF13* was expressed by myeloid lineage cells, *CCR1* was expressed by macrophages which appear to have an M2 phenotype, and *TNFSFR10A* is expressed by B cells (data not shown). A similar approach could be employed for the remaining interactions to further resolve the interacting cells of interest.

### **Tubular cells and keratinocytes from patients with proliferative histologic classes compared to membranous class upregulate TNF and type-I IFN response pathways**

The molecular basis for different histopathologies in LN is not completely understood. To determine if there are specific pathways involved in each class, we performed differential expression on tubular cells from patients with proliferative LN (class III or class IV) and those with membranous disease (class V). This analysis excluded patients with mixed class III/V or IV/V disease. Pathway enrichment analysis of the upregulated genes in proliferative class disease revealed increased type-I IFN and TNF family signaling compared with tubular



cells from membranous class (Fig. 7). Keratinocytes from patients with proliferative disease also showed an upregulation of several pathways, including type-I IFN signaling and antigen presentation, compared with keratinocytes from membranous disease (Fig. 7). The full list of differentially expressed genes can be found in Supplementary Table 4. Thus, this analysis showed that patients with proliferative class LN upregulated several immune related pathways in both their tubular cells and keratinocytes which differentiated them from patients with membranous class of LN.

### **Tubular cells from patients with mixed histologic classes upregulate distinct expression profiles as compared to membranous or proliferative nephritis**

We performed differential expression analysis between membranous and mixed class LN, and between proliferative and mixed class LN. There were several differentially expressed pathways between membranous and mixed class LN, such as IL-1 signaling which was upregulated in mixed versus membranous (Fig. 8a). While there were no upregulated pathways in mixed versus proliferative, there were several upregulated pathways in proliferative disease including TNFR1 signaling (Fig. 8b). Thus, while mixed class histology is commonly considered a simple combination of class III or IV with class V, based on our findings it is possible that it represents a distinct class of disease with different pathological pathways contributing to the histologic phenotype.

## **Discussion**

In this study, scRNA-seq applied to renal and skin biopsies from patients with LN and healthy control subjects identified clinically relevant signatures associated with disease. The focus and strength of our analysis were the epithelial cells, which accounted for the majority of cells in the kidney. Their pathological potential to promote and contribute to fibrosis (resulting in poor prognosis) and their utility as a source of biomarkers were explored. Although patient numbers were relatively small when compared to GWAS or other such large-scale evaluations, this study provides the largest scRNA-seq database to date, covering all major classes of LN, and further including healthy donor kidney reference data.

As previously reported, we discovered an elevated IFN response signature in keratinocytes from patients with LN compared to healthy control subjects<sup>9</sup>, indicative of a systemic response to IFN. Furthermore, tubular cells from renal biopsies obtained from patients with LN compared to healthy control subjects also showed an elevated IFN response signature which correlated with keratinocyte expression. We examined other cell types such as fibroblasts, mesangial cells, and endothelial cells, and although the trend was the same, the cell numbers were not sufficient to draw meaningful conclusions. While an IFN response signature may also have been evident by bulk sequencing, identification of this signature specifically in tubular cells required scRNA-seq. Accordingly, this study unambiguously identified that tubular cells per se upregulate IFN signature genes, and not just infiltrating cells that could display a large enough IFN response to be captured by bulk sequencing despite their relative rarity. A signature present and visible in the dominant cell type of the biopsy justifies applying bulk RNA-seq or RT-PCR as a routine diagnostic in the clinical setting without the need for complicated cell sorting. Moreover, although we have no

evidence of causation, the tubular IFN response score at the time of biopsy predicted patient response to treatment 6 months after biopsy. The keratinocyte IFN response score did not significantly predict response to treatment. This may be explained by a weaker signal than that observed in the tubular cells and thus requiring larger numbers to reach significance, and/or relate to the tissue contextual dependency of downstream expression of IFN stimulated genes.

In addition to the IFN signature, pathways associated with ECM proteins and ECM-interacting proteins indicative of a fibrotic response were significantly upregulated in the tubular cells of non-responder patients. Tubular ECM protein expression has been linked to tubular EMT, a process whereby tubular cells differentiate into mesenchymal cells and begin secreting large amounts of ECM proteins<sup>18,20</sup>. Although still a topic of debate, EMT is linked to increased interstitial fibrosis and by extension poor prognosis<sup>28–30</sup>. We did not detect putative markers of EMT other than *COL1A1* and *COL1A2*, suggesting that these cells may be early in this differentiation pathway<sup>29</sup> or the findings represent a separate fibrotic pathway, such as tubular epithelial hypertrophy<sup>21</sup>. Remarkably, this fibrotic gene expression was independent of histologic evidence of fibrosis, thus providing an important adjunct to clinical evaluation. Including such a diagnostic at the time of biopsy could potentially guide the use of more aggressive therapy to control fibrotic scar formation leading to organ failure.

Pathways indicative of a fibrotic signature in the tubular cells were also significantly upregulated in the keratinocytes of non-responders, and furthermore there was a correlation in individual patients between fibrotic scores in tubular cells and keratinocytes. Although the 4-gene model generated in tubular cells to predict response to treatment was not significant in the keratinocytes, this may change with higher throughput and more sensitive technology.

By investigating the receptor-ligand interactions among cell types in the skin and kidney, we identified several putative signaling interactions. For instance, interactions were identified between infiltrating leukocytes and tubular cells through an FGF receptor, the latter known to be involved in fibrotic processes and likely responsible for the upregulation of genes encoding ECM and ECM-interacting proteins in the tubular cells of non-responder patients. Genes encoding FGF receptors were highly expressed in all resident kidney cells, including fibroblasts, endothelial cells, and mesangial cells. Additionally, chemokines produced by resident renal cells including tubular cells, endothelial cells, and fibroblasts may be involved in the recruitment of inflammatory cells into the kidney. Further validating these interactions would be important to address in future in vitro and animal studies. It is anticipated that exploring such interactions will be critical to developing novel therapeutic targets, which can be employed for disease-state-specific treatment based on molecular diagnosis.

Examining skin and kidney epithelial cells revealed molecular signatures which differentiated histological classes of LN. Pathway enrichment analysis and differential expression in tubular cells from proliferative versus either mixed or membranous class LN indicated an upregulation of genes involved in TNF family signaling. Furthermore, many other pathways identified in both skin and kidney differentiated the classes, including type-I IFN signaling. Consistent with these results, a previous study using microarray technology

and antibody staining demonstrated an IFN signature in class IV but not class I LN tubular cells<sup>31</sup>. Our dataset provides an opportunity to explore these pathways to both improve diagnostics and inform the development of novel and potentially class specific therapies.

While our previous scRNA-seq study of renal biopsies using the Fluidigm 96-well platform yielded most of the dominant renal cell types, glomerular cells were absent<sup>9</sup>. Using the 800-well platform markedly increased cell capture, and importantly allowed for the capture and identification of mesangial cell profiles from both healthy control subjects and patients with LN. Podocytes, however, were not captured; a further increase in throughput by using next-generation droplet-based microfluidics, recently shown to provide increased resolution of cell types and cell states, may prove necessary to capture this rarer population of cells. These latter studies, however, were not performed using very limited biopsy tissue (2–10 mm needle core pieces) obtained for clinical decision making, but rather on an excess of tissue from whole organs or surgically resected tissues<sup>13,32–34</sup>. While we could resolve mesangial cells, endothelial cells, and fibroblasts, their relatively low abundance limited their ability to discriminate between patients and patient groups<sup>35</sup>. Consequently, our study focused on the dominant epithelial cells based on clusters formed of cells from many patients. Since batch correction was not performed due to the inherent difficulty in defining a batch when every patient's samples are processed separately, we performed differential expression at the cell population level generated through “course” clustering, which by design minimized cell cluster formation based on individual patients. Approaches to increase the number of each cell type captured will be needed to evaluate the minor cell populations of the skin and kidney.

In summary, we have shown that scRNA-seq is feasible and informative in the study of LN, despite the marked complexity and heterogeneity of the disease. Our findings support the utility of skin biopsies as a potential adjunct in the management of LN since repeat biopsies can be performed regularly. scRNA-seq of LN tissues revealed molecular signatures clinically relevant to prognosis, which could be used to meaningfully augment the current standard of care and better stratify patients. Moreover, these molecular signatures also begin to reveal processes which may underlie the histologic heterogeneity of LN.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This work was supported by the Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus Network. AMP is a public-private partnership (AbbVie, Arthritis Foundation, Bristol-Myers Squibb, Foundation for the National Institutes of Health, Lupus Foundation of America, Lupus Research Alliance, Merck Sharp & Dohme, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer, Rheumatology Research Foundation, Sanofi, and Takeda Pharmaceuticals) created to develop new ways of identifying and validating promising biological targets for diagnostics and drug development. Funding was provided through grants from the National Institutes of Health (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-AR067685, UH2-AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-AR067694, and UM2-AR067678). We thank the Rockefeller University Genomics Resource Center for providing access to the Fluidigm C1 system and Illumina sequencing.

## References

1. Tsokos GC Systemic Lupus Erythematosus. *N. Engl. J. Med* 365, 2110–2121 (2011). [PubMed: 22129255]
2. Mohan C & Putterman C Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat. Rev. Nephrol* 11, 329–341 (2015). [PubMed: 25825084]
3. Weening JJ et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J. Am. Soc. Nephrol. JASN* 15, 241–250 (2004). [PubMed: 14747370]
4. Yu F et al. Tubulointerstitial lesions of patients with lupus nephritis classified by the 2003 International Society of Nephrology and Renal Pathology Society system. *Kidney Int.* 77, 820–829 (2010). [PubMed: 20182417]
5. Hsieh C et al. Predicting outcomes of lupus nephritis with tubulointerstitial inflammation and scarring. *Arthritis Care Res.* 63, 865–874 (2011).
6. Alsuwaida AO Interstitial inflammation and long-term renal outcomes in lupus nephritis. *Lupus* 22, 1446–1454 (2013). [PubMed: 24097970]
7. Misra R & Gupta R Biomarkers in lupus nephritis. *Int. J. Rheum. Dis* 18, 219–232 (2015). [PubMed: 25884459]
8. Reich A, Marcinow K & Bialynicki-Birula R The lupus band test in systemic lupus erythematosus patients. *Ther. Clin. Risk Manag.* 7, 27–32 (2011). [PubMed: 21339940]
9. Der E et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight* 2,
10. Ofengeim D, Giagtzoglou N, Huh D, Zou C & Yuan J Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. *Trends Mol. Med* 23, 563–576 (2017). [PubMed: 28501348]
11. Satija R, Farrell JA, Gennert D, Schier AF & Regev A Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol* 33, 495–502 (2015). [PubMed: 25867923]
12. Kim K-T et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* 17, 80 (2016). [PubMed: 27139883]
13. Young MD et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361, 594–599 (2018). [PubMed: 30093597]
14. Park J et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* eaar2131 (2018). doi:10.1126/science.aar2131
15. Schitteck B et al. Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nat. Immunol* 2, 1133–1137 (2001). [PubMed: 11694882]
16. Du J et al. MLANA/MART1 and SILV/PMEL17/GP100 Are Transcriptionally Regulated by MITF in Melanocytes and Melanoma. *Am. J. Pathol* 163, 333–343 (2003). [PubMed: 12819038]
17. Elkon KB & Stone VV Type I Interferon and Systemic Lupus Erythematosus. *J. Interferon Cytokine Res.* 31, 803–812 (2011). [PubMed: 21859344]
18. Lan HY Tubular epithelial-myofibroblast transdifferentiation mechanisms in proximal tubule cells. *Curr. Opin. Nephrol. Hypertens* 12, 25–29 (2003). [PubMed: 12496662]
19. Ng YY et al. Tubular epithelial-myofibroblast transdifferentiation in progressive tubulointerstitial fibrosis in 5/6 nephrectomized rats. *Kidney Int.* 54, 864–876 (1998). [PubMed: 9734611]
20. Lamouille S, Xu J & Derynck R Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15, 178–196 (2014). [PubMed: 24556840]
21. Wolf G & Ziyadeh FN Renal tubular hypertrophy induced by angiotensin II. *Semin. Nephrol* 17, 448–454 (1997). [PubMed: 9316213]
22. Zhang X et al. TIMP-1 promotes age-related renal fibrosis through upregulating ICAM-1 in human TIMP-1 transgenic mice. *J. Gerontol. A. Biol. Sci. Med. Sci* 61, 1130–1143 (2006). [PubMed: 17167153]
23. Ling XB et al. Integrative Urinary Peptidomics in Renal Transplantation Identifies Biomarkers for Acute Rejection. *J. Am. Soc. Nephrol* 21, 646–653 (2010). [PubMed: 20150539]
24. Su Z et al. Excessive activation of the alternative complement pathway in autosomal dominant polycystic kidney disease. *J. Intern. Med* 276, 470–485 (2014). [PubMed: 24494798]

25. Parikh SV et al. Molecular imaging of the kidney in lupus nephritis to characterize response to treatment. *Transl. Res* 182, 1–13 (2017). [PubMed: 27842222]
26. Strutz F et al. Basic fibroblast growth factor expression is increased in human renal fibrogenesis and may mediate autocrine fibroblast proliferation. *Kidney Int.* 57, 1521–1538 (2000). [PubMed: 10760088]
27. Smith ER, Tan S-J, Holt SG & Hewitson TD FGF23 is synthesised locally by renal tubules and activates injury-primed fibroblasts. *Sci. Rep* 7, 3345 (2017). [PubMed: 28611350]
28. Meran S & Steadman R Fibroblasts and myofibroblasts in renal fibrosis. *Int. J. Exp. Pathol* 92, 158–167 (2011). [PubMed: 21355940]
29. Yang J & Liu Y Dissection of key events in tubular epithelial to myofibroblast transition and its implications in renal interstitial fibrosis. *Am. J. Pathol* 159, 1465–1475 (2001). [PubMed: 11583974]
30. Yang J & Liu Y Blockage of tubular epithelial to myofibroblast transition by hepatocyte growth factor prevents renal interstitial fibrosis. *J. Am. Soc. Nephrol. JASN* 13, 96–107 (2002). [PubMed: 11752026]
31. Castellano G et al. Local synthesis of interferon-alpha in lupus nephritis is associated with type I interferons signature and LMP7 induction in renal tubular epithelial cells. *Arthritis Res. Ther* 17, (2015).
32. Stephenson W et al. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat. Commun* 9, (2018).
33. MacParland SA et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun* 9, (2018).
34. Vento-Tormo R et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* 563, 347 (2018). [PubMed: 30429548]
35. Austin HA, Boumpas DT, Vaughan EM & Balow JE Predicting renal outcomes in severe lupus nephritis: contributions of clinical and histologic data. *Kidney Int.* 45, 544–550 (1994). [PubMed: 8164443]
36. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* 17, 10–12 (2011).
37. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl* 29, 15–21 (2013).
38. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl* 30, 923–930 (2014).
39. Macosko EZ et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214 (2015). [PubMed: 26000488]
40. Pavli ev M et al. Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* 27, 349–361 (2017). [PubMed: 28174237]
41. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, (2014).
42. Kuleshov MV et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–97 (2016). [PubMed: 27141961]
43. Chen EY et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128 (2013). [PubMed: 23586463]
44. Croft D et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* 42, D472–477 (2014). [PubMed: 24243840]
45. Fabregat A et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46, D649–D655 (2018). [PubMed: 29145629]
46. Kanehisa M & Goto S KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30 (2000). [PubMed: 10592173]
47. Kanehisa M, Furumichi M, Tanabe M, Sato Y & Morishima K KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361 (2017). [PubMed: 27899662]

48. Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–462 (2016). [PubMed: 26476454]

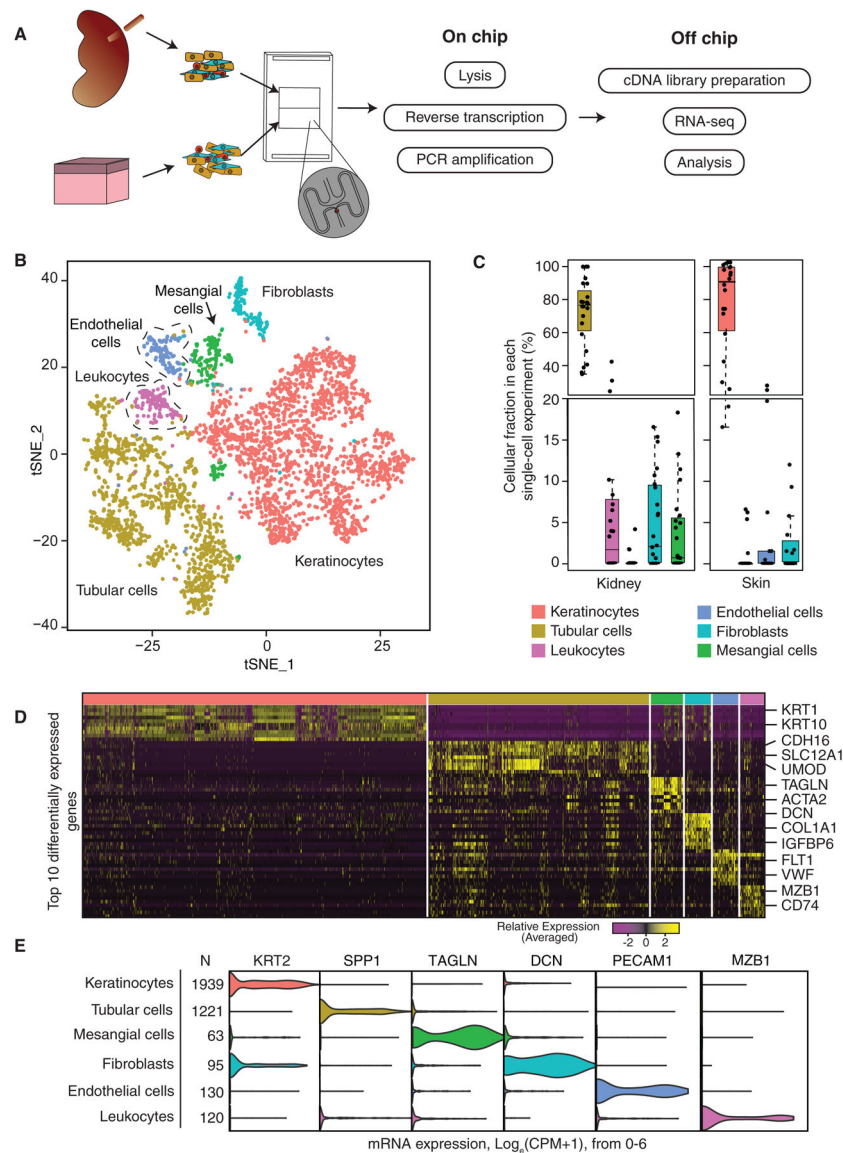
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



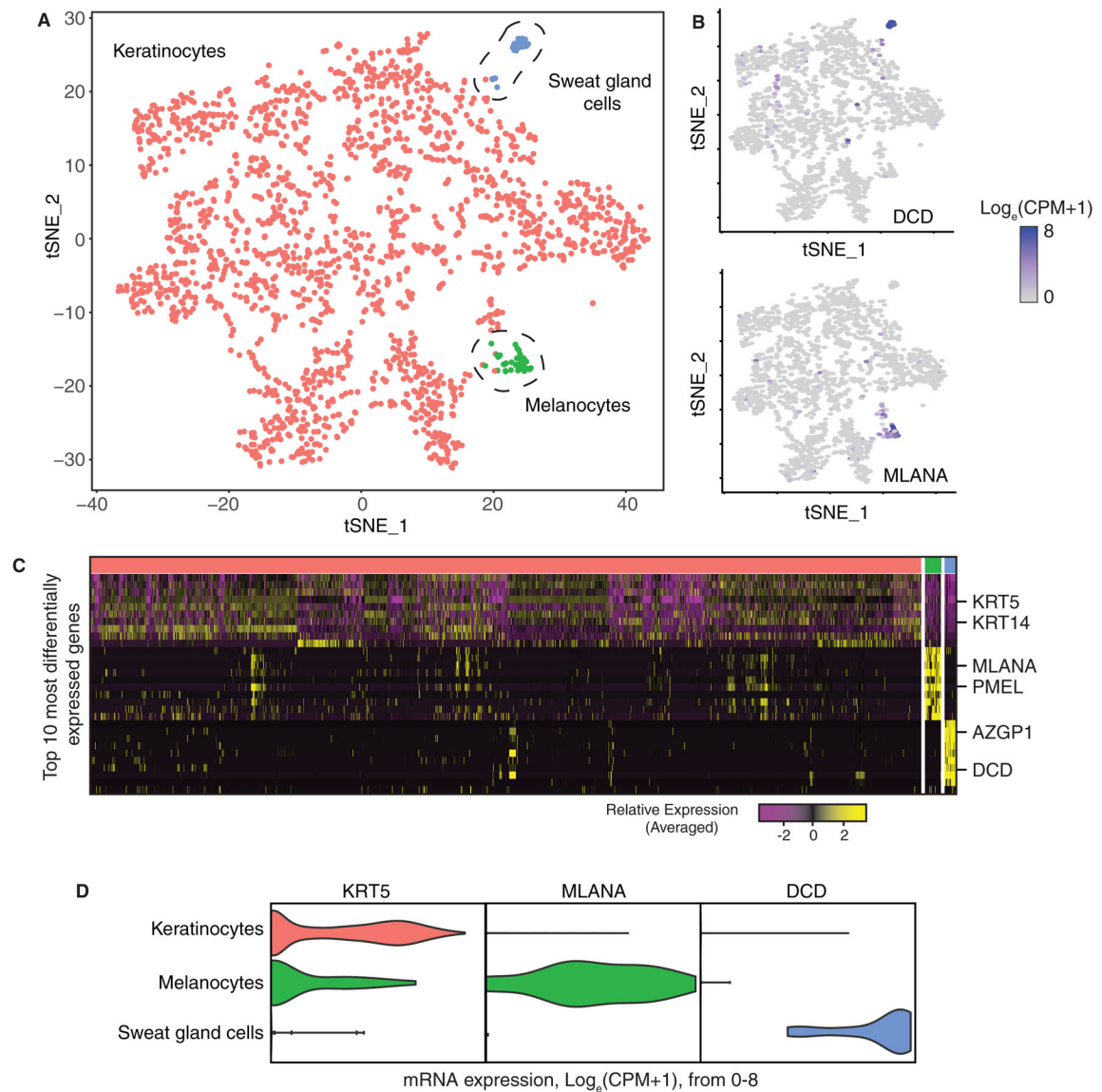


**Figure 1. Cell lineage determination by dimensionality reduction analysis.**

**a.** Schematic of the scRNA-seq pipeline. Skin (n = 17) and kidney (n = 21) samples from patients with LN or healthy control subjects (n = 3) were collected at the time of clinically indicated renal biopsy or live kidney donation, respectively. Skin and kidney biopsies were enzymatically disaggregated into single cell suspensions and loaded onto a microfluidic device. **b.** t-Distributed Stochastic Neighbor Embedding (tSNE) clustering of 4,019 single cells. PCA identified six major clusters of cells from both skin and kidney biopsies. Cells are color-coded by an algorithm for determining expression clusters and cell types. **c.** Box plot of the percent contribution of each cluster from skin and kidney biopsies. Boxes are colored by cluster corresponding to Fig. 1b. The boxes indicate the first quartile, median, and third quartile. Whiskers indicate the highest and lowest values. Points were drawn as outliers if they were more than 1.5 times the inter quartile range. **d.** Heatmap of the top 10 most differentially expressed genes in each cluster to identify mutually exclusive gene sets, which

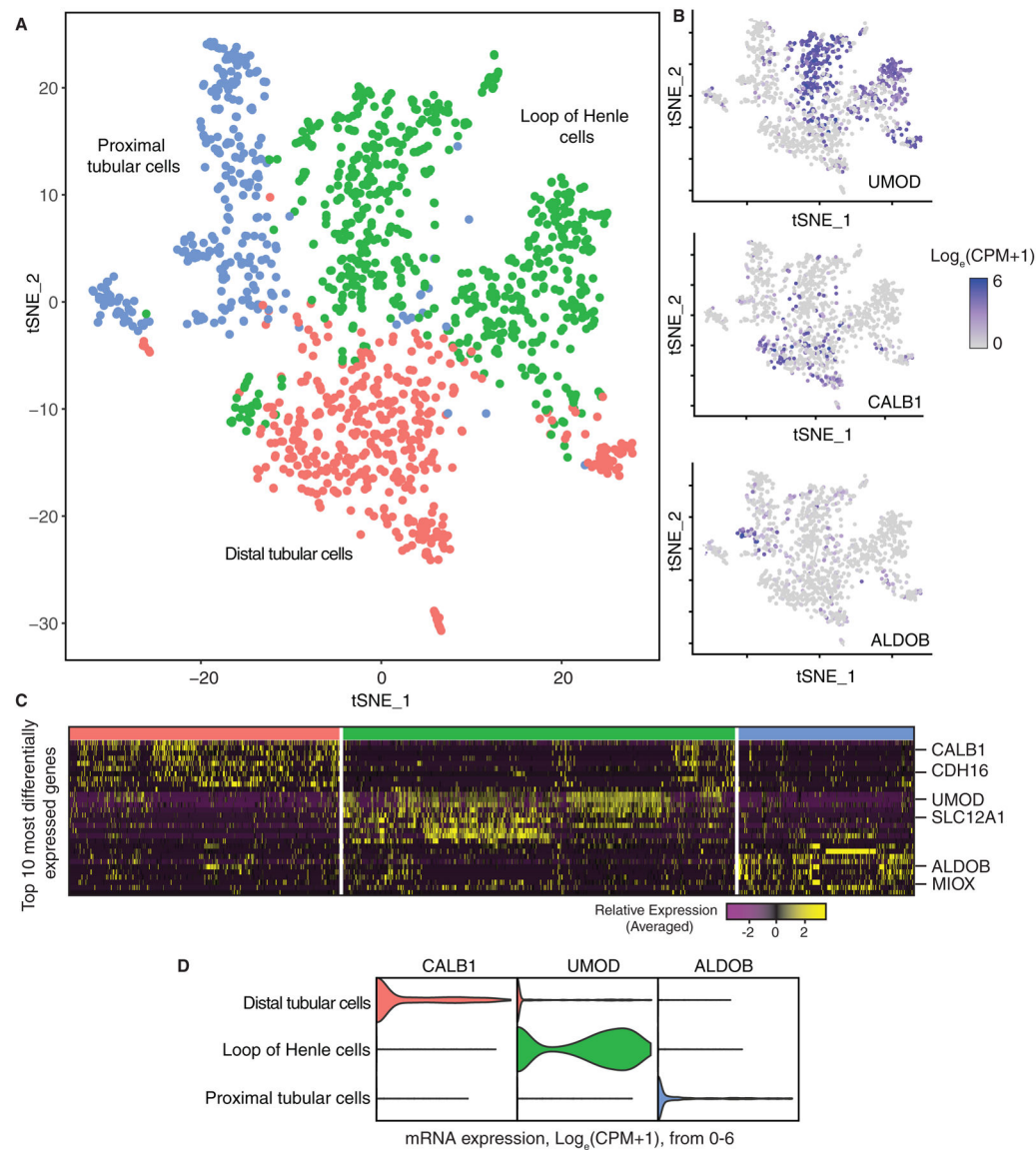


were then used to determine the cell lineage of each cluster. Each row is a differentially expressed gene and each column is a single-cell organized by cluster identity. Transcript abundance ranges from low (purple) to high (yellow). **e.** Violin plot of selected markers indicating the expression level of canonical markers within each cluster. Violin plots are colored by cell type and width represents the percent of cells expressing the marker at a given level in  $\text{Log}_e(\text{CPM}+1)$  from 0 to 6.

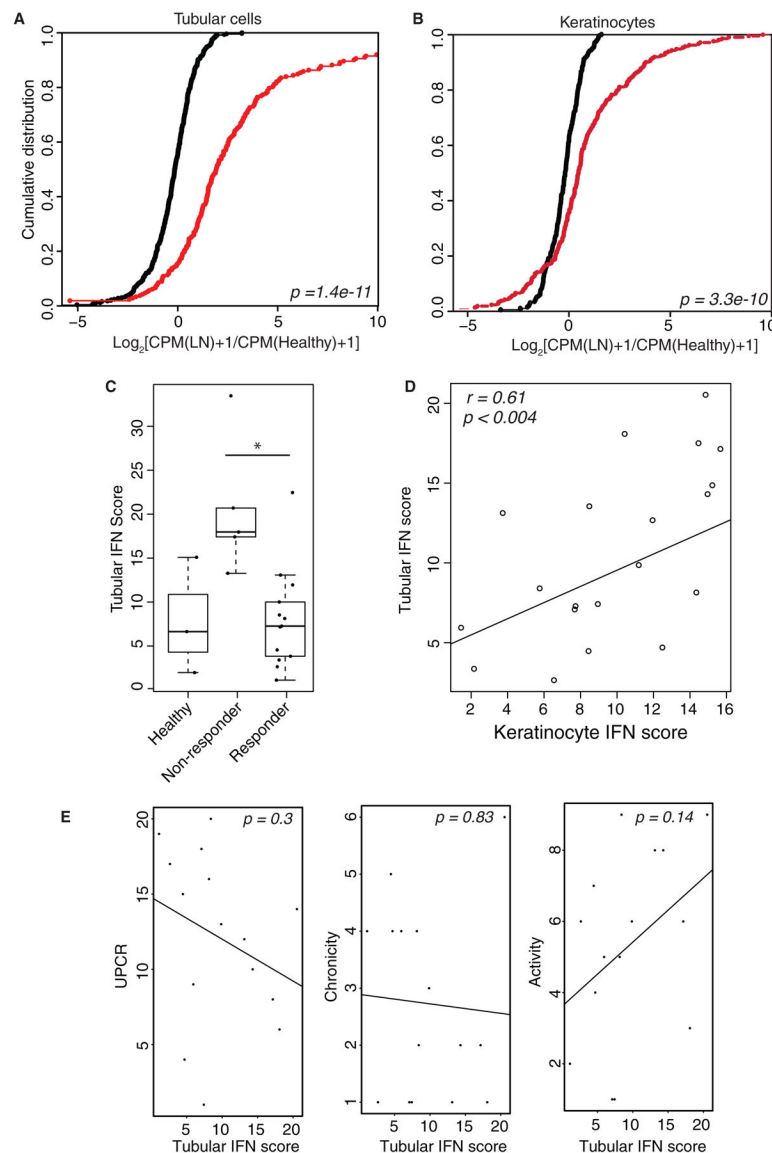


**Figure 2. Subclustering of keratinocytes reveals two rare skin-specific cell types.**

**a.** tSNE plotting of 1,939 keratinocytes identified in initial clustering analysis colored by cluster identifying algorithm with cell type labels next to each cluster. **b.** Expression of *DCD* and *MLANA*, markers of the new clusters, within the tSNE plot ( $n = 1,939$  cells) from low expression (grey) to high expression (dark blue). **c.** Heatmap of the differentially expressed genes between each identified cluster. **d.** Violin plot of *MLANA* and *DCD*, markers of melanocytes ( $n = 33$ ) and sweat gland cells ( $n = 29$ ), respectively. Violin plots are colored by cell type and width represents the percent of cells expressing the marker at a given level in  $\text{Log}_e(\text{CPM}+1)$  from 0 to 8.

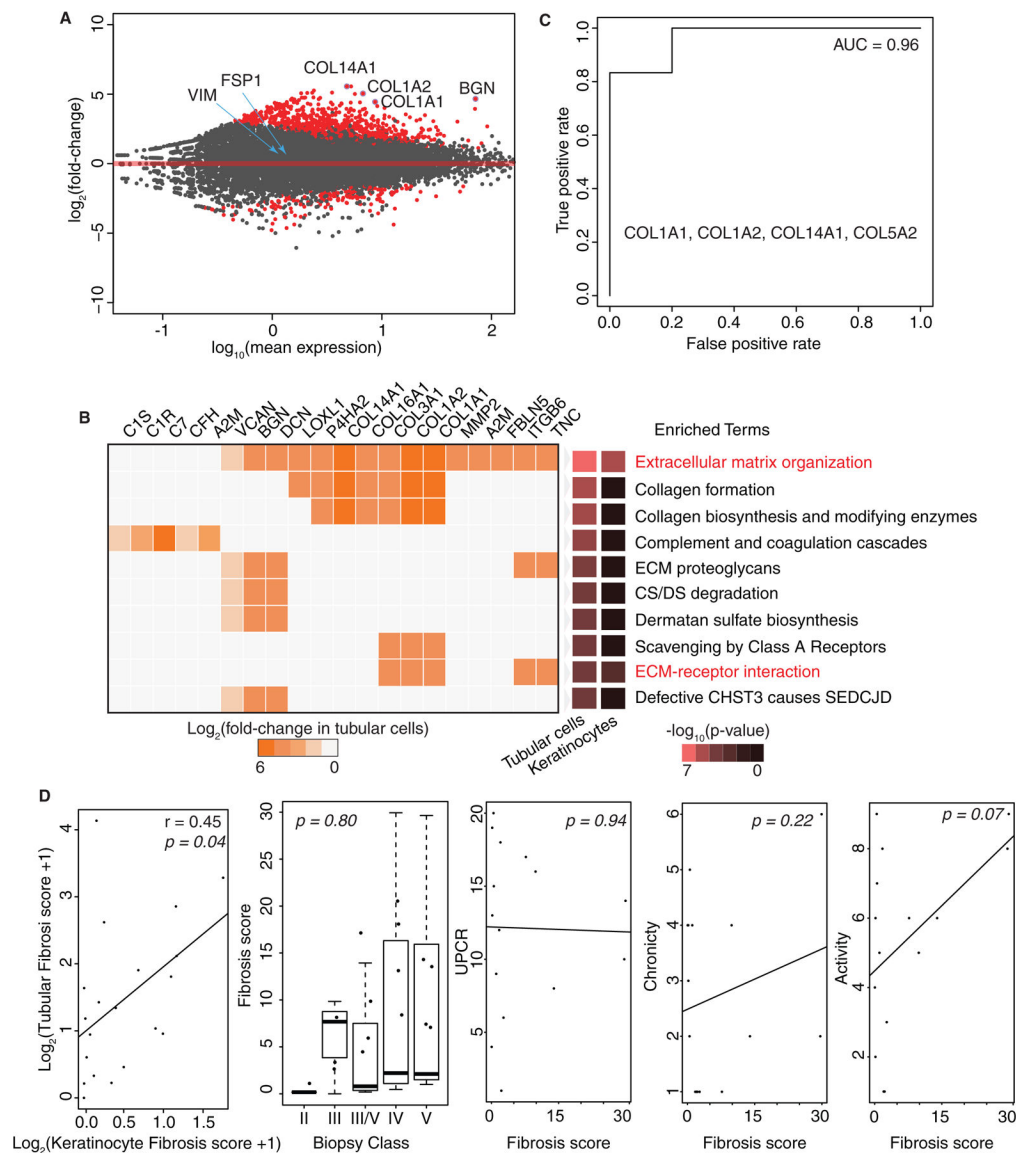


**Figure 3. Subclustering of tubular cells identifies major tubular cell subtypes of the nephron.**  
**a.** tSNE plotting of 1,221 tubular cells identified by initial clustering analysis. Three clusters of tubular cells are identified and colored by clustering algorithm with labels of putative cluster identity indicated next to each cluster. **b.** Expression of established tubular subtype markers within the tSNE plot ( $n = 1,221$  cells) from low expression (grey) to high expression (dark blue). **c.** Heatmap of the top 10 most differentially expressed genes between each cluster. **d.** Violin plots of *UMOD*, *CALB1*, and *ALDOB* which are canonical markers of Loop of Henle ( $n = 581$ ), distal tubular ( $n = 394$ ), and proximal tubular cells ( $n = 246$ ), respectively. Violin plots are colored by cell type and width represents the percent of cells expressing the marker at a given level in  $\text{Log}_e(\text{CPM}+1)$  from 0 to 6.



**Figure 4. IFN response signature differentiates patients with LN from healthy control subjects and response to treatment.**

**a.** Cumulative distribution function (CDF) of the ratio of expression of 212 IFN responsive genes (red line) or ubiquitously expressed genes (black line) in both tubular cells (n = 1,112 patient cells and n = 109 healthy control cells,  $p = 1.4 \times 10^{-11}$ ) and **b.** keratinocytes (n = 1,766 patient cells and n = 173 healthy control cells,  $p = 3.3 \times 10^{-10}$ ) compared using a two-tailed Wilcoxon signed rank test. **c.** Boxplot of IFN response scores in healthy control subjects (n = 3), patients who responded to treatment (n = 13), and patients who did not respond to treatment (n = 5) compared with a two-tailed Student's t-test ( $p = 0.0003$ ,  $t = 4.7234$ ). The boxes indicate the first quartile, median, and third quartile. Whiskers indicate the highest and lowest values. Points were drawn as outliers if they were more than 1.5 times the inter quartile range. **d.** Pearson's correlation between tubular and keratinocyte IFN response scores per patient (n = 20,  $r = 0.61$ ,  $p = 0.004$ ). **e.** Pearson's correlations between IFN response scores and UPCR ( $p = 0.3$ ), Chronicity ( $p = 0.83$ ), and Activity indices ( $p = 0.14$ ).

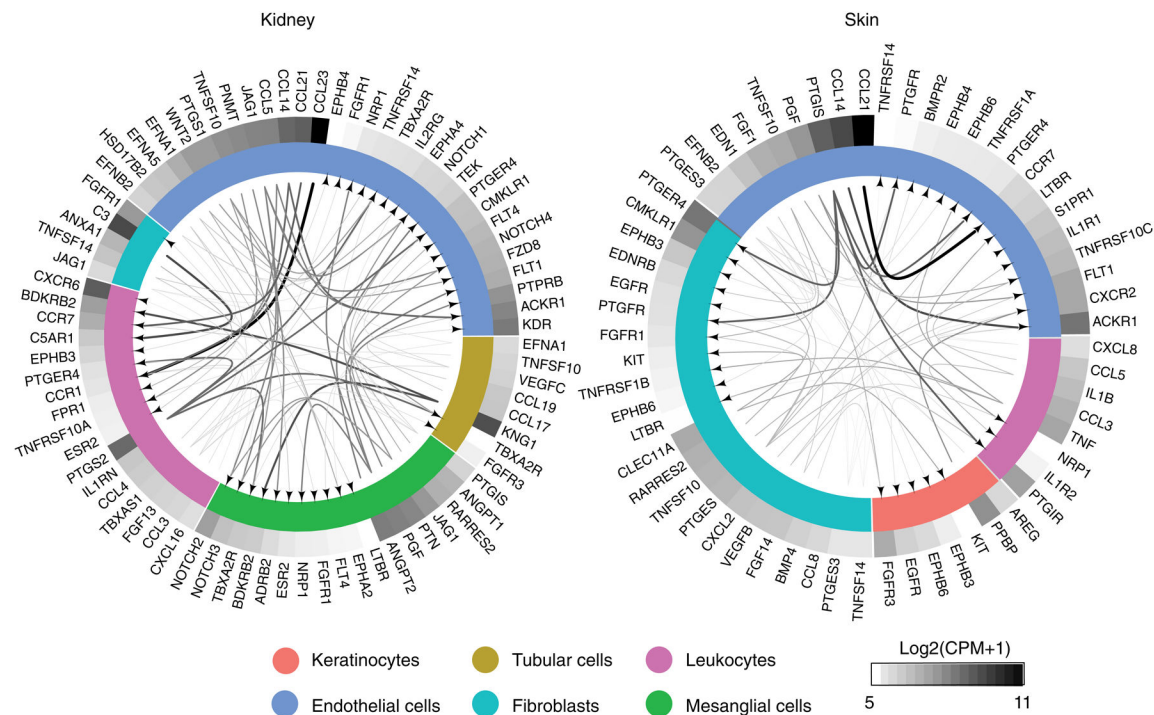


**Figure 5. A fibrotic gene signature as a potential prognostic marker for patients non-responsive to treatment.**

**a.** MA plot of differential expression analysis performed between tubular cells of patients responsive ( $n = 13$ ) or non-responsive to treatment ( $n = 5$ ). Significantly differentially expressed genes determined by the Wald test corrected for multiple comparisons, are colored in red. **b.** Pathway enrichment analysis of genes identified as upregulated in patients non-responsive to treatment ( $n = 13$ ) in Figure 5A.  $-\log_{10}(\text{p-value})$  determined by gene ontology fuzzy-enrichment analysis of each pathway is shown for both keratinocytes and tubular cells colored from least significant (black) to most significant (red).  $\log_2$  fold change in gene expression between patients non-responsive to treatment ( $n = 13$ ) compared with patients responsive to treatment ( $n = 5$ ) in each pathway are indicated for tubular cells from smallest (grey) to highest (orange). **c.** Receiver operating characteristic curve of the logistic regression equation of differentially expressed fibrotic genes, *COL1A2*, *COL1A1*, *COL14A1*, *COL5A2*, with area under the curve (AUC) indicated ( $n = 18$  patients). **d.**

Pearson's Correlations between the  $\log_2$  transformed tubular and keratinocyte fibrosis scores ( $n = 20$ ,  $r = 0.45$ ,  $p = 0.04$ ), the tubular fibrosis score by biopsy class ( $p = 0.80$ ,  $F = 0.334$ ), and between tubular fibrosis scores and UPCR ( $p = 0.94$ ), Chronicity ( $p = 0.22$ ), and Activity indices ( $p = 0.07$ ). The boxes indicate the first quartile, median, and third quartile. Whiskers indicate the highest and lowest values.



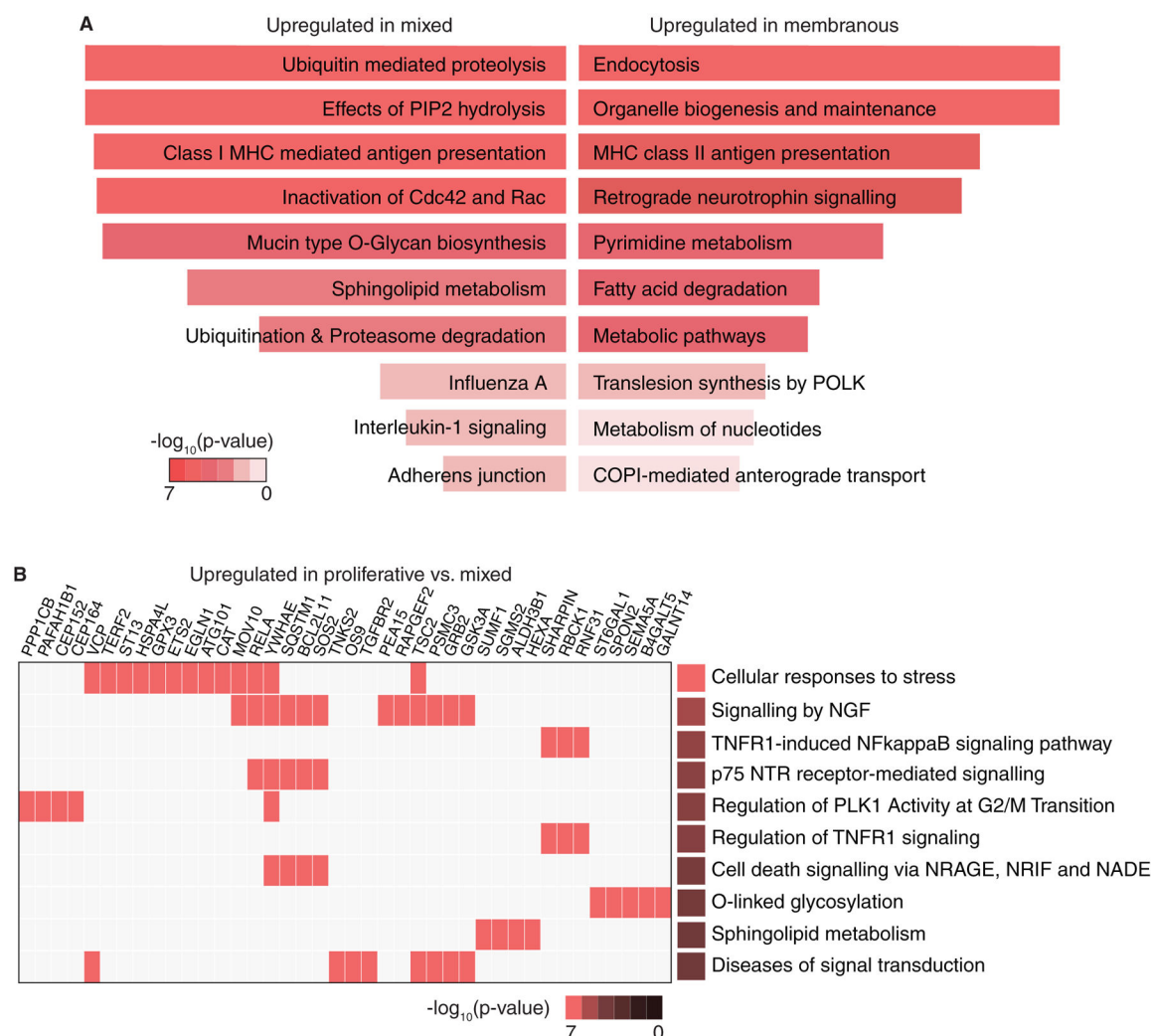


**Figure 6. Putative receptor-ligand interactions between kidney and skin cells.** Lines represent interactions between cell types in the skin and the kidney, and are shaded according to the ligand expression as detailed in the scale bar. Lines originate at the ligand and connect to its receptor as indicated by the arrowhead. Each cell type is color coded and represented by that color in each organ. Only the top expressed receptors and ligands with expression above 45 CPM and 65 CPM, respectively, within each cell type are shown. Receptors and ligand are arranged by expression strength clockwise from lowest expression to highest. Ligands or receptors without a cognate pair were excluded from visualization. Only patients with LN (n = 18) were included in this analysis.





**Figure 7. Differential expression and pathway enrichment analysis of tubular cells and keratinocytes between membranous and proliferative LN.** Significantly enriched pathways in both tubular cells and keratinocytes in membranous (n = 6) and proliferative (n = 8) LN are indicated. Mixed class III/V or IV/V were excluded from this analysis. Color intensity and length of bar indicates higher-Log<sub>10</sub>(p-value) determined by gene ontology fuzzy-enrichment analysis from least significant (white) to most significant (red).



**Figure 8. Differential expression and pathway enrichment analysis of tubular cells between membranous or proliferative LN and mixed class disease.**

**a.** Differential expression analysis and pathway enrichment of mixed class nephritis (class III/V and class IV/V,  $n = 5$ ) vs. membranous (class V,  $n = 6$ ). Color intensity and length of the bar indicate higher- $\text{Log}_{10}(\text{p-value})$  determined using gene ontology fuzzy-enrichment analysis from least significant (white) to most significant (red). **b.** Upregulated pathways and genes in proliferative nephritis (class III and class IV,  $n = 8$ ) vs. mixed class nephritis. -  $\text{Log}_{10}(\text{p-value})$  determined using gene ontology fuzzy-enrichment analysis of each pathway is shown and colored from least significant (black) to most significant (red).